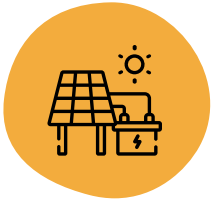
Renewable Energy Forecasting

Jocelyn Zhao, Cecilia Dong, Charlene Wei

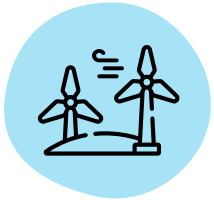
Agenda

01



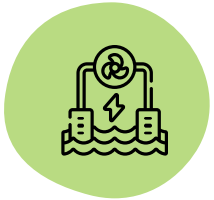
Project Objective

02



Data Preparation

03

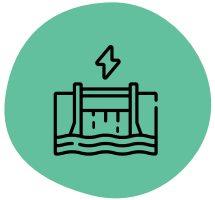


Exploratory Analysis

04

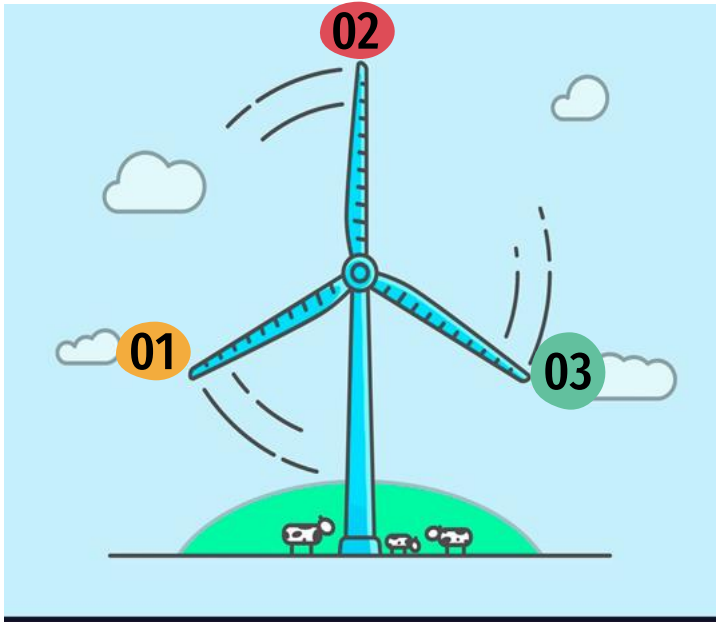
Models Processing

05



Final Model & Forecast

Project Objective



Background Information

01

As wind farms becomes more prevalent, the wind power forecasting become more crucial

Company Needs

02

The utility firms need precise estimates of the electricity generated by wind turbines to meet demand

Goal

03

Create a daily forecast for Active Power Generation for the turbine for the next 5 day period

Data Preparation - changing time index

```
turbine <- turbine %>%
    na_interpolation()%>%
  group_by(year, month, day) %>%
  summarize(active_power = sum(active_power, na.rm = TRUE),
            ambient_temperature = mean(ambient_temperature, na.rm = TRUE),
            wind_direction = mean(wind_direction, na.rm = TRUE),
            wind_speed = mean(wind_speed, na.rm = TRUE))
```

year	month	day	date
2017	12	31	2017-12-31 00:00:00
2017	12	31	2017-12-31 00:10:00
2017	12	31	2017-12-31 00:20:00
2017	12	31	2017-12-31 00:30:00
2017	12	31	2017-12-31 00:40:00
2017	12	31	2017-12-31 00:50:00
2017	12	31	2017-12-31 01:00:00

year	month	day
2017	12	31
2018	1	1
2018	1	2
2018	1	3
2018	1	4

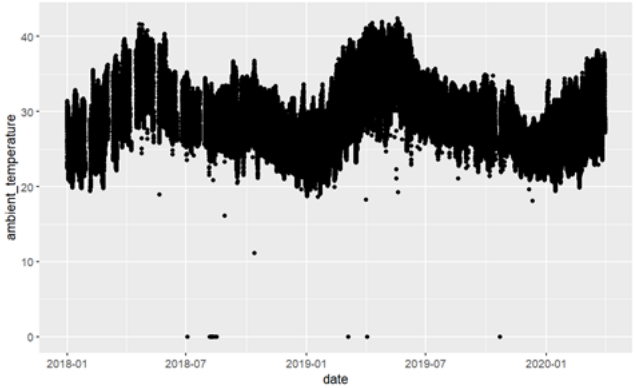
Data Preparation - missing value

	skim_variable <chr>	n_missing <int>	c
1	year	0	
2	month	0	
3	day	0	
4	active_power	23330	
5	ambient_temperature	24263	
6	wind_direction	45802	
7	wind_speed	23485	

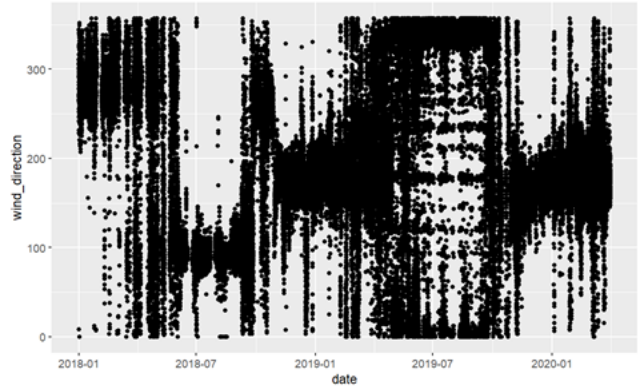
year <dbl>	month <dbl>	n_missing <int>	co
2019	10	0	
2019	11	0	
2019	12	0	
2020	1	0	
2020	2	0	
2020	3	0	
2017	12	0	
2018	1	0	
2018	2	0	
2018	3	0	

Exploratory Analysis - explanatory variables

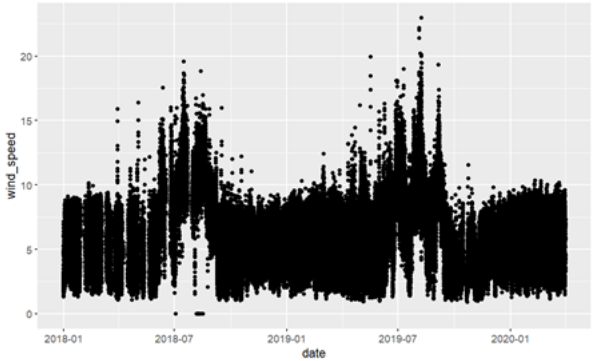
Ambient Temperature



Wild Direction



Wind Speed

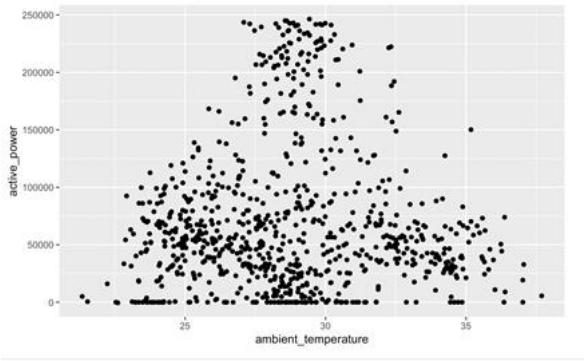


Plotting variables with time

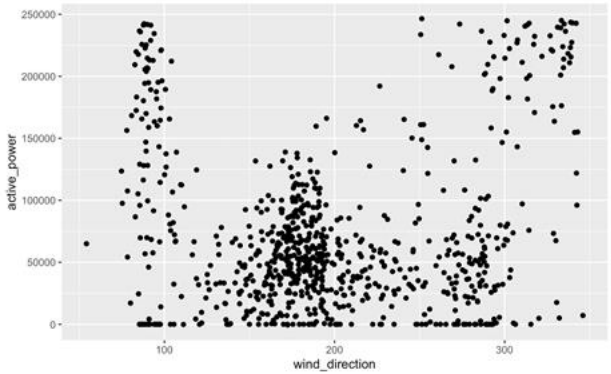
No Event Variables

Exploratory Analysis - explanatory variables

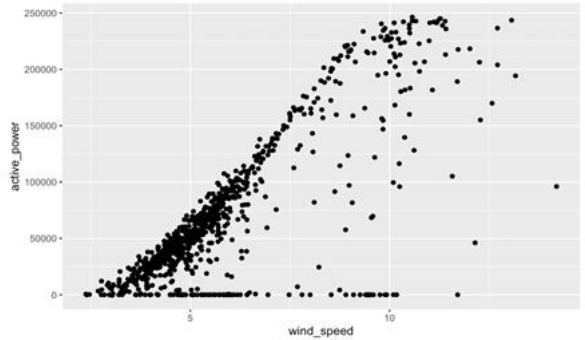
Ambient Temperature



Wild Direction



Wind Speed

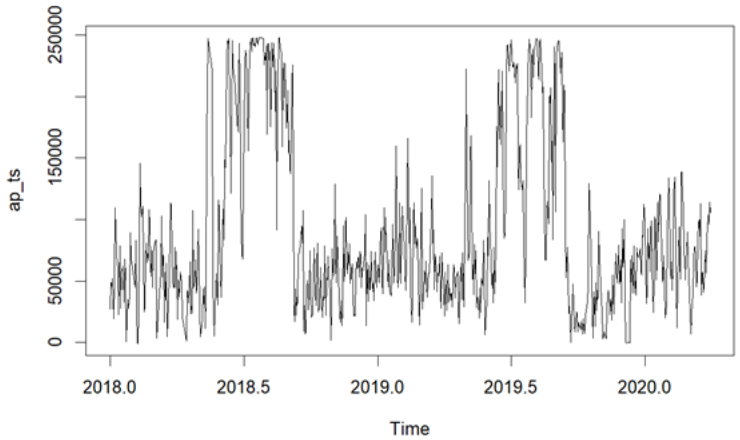


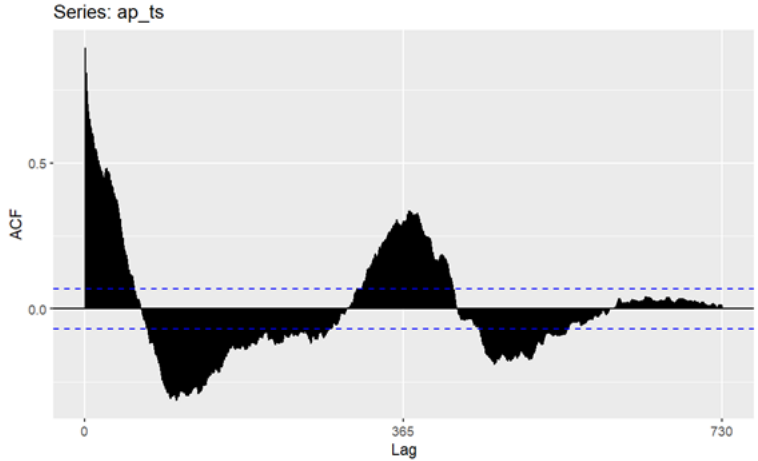
Plotting variables with response variable

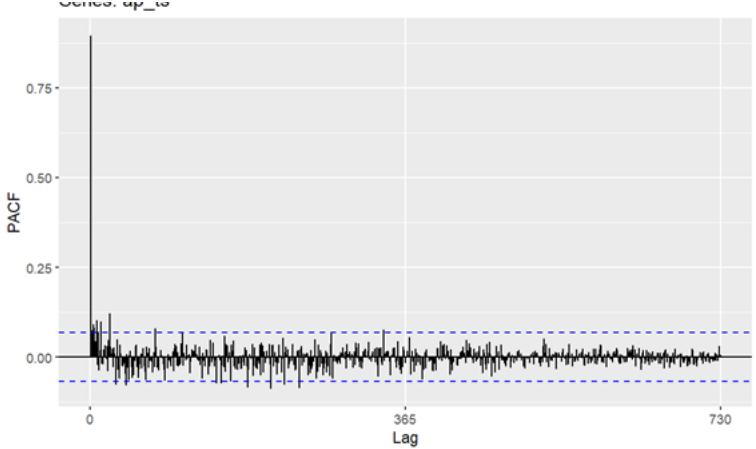
There is correlation between wind speed and active power

Exploratory Analysis - time series

- **Time series:** values around May each year are higher than other values
- **ACF:** decays
- **PACF**: cut off quickly







Models Processing

1. **Exploring linear regression first**

   (all variables are significant at significance level 0.05)

	Estimate	SE	t.value	p.value
intercept	-102582.4490	8210.0225	-12.4948	0.0000
ambient_temperature	610.2466	261.3685	2.3348	0.0198
wind_direction	-42.5943	12.7461	-3.3418	0.0009
wind_speed	30919.5702	408.8773	75.6207	0.0000

1. **Auto.arima > arima(1,0,1)**

- RMSE: 16046.99
- Test for white noise (p-value = 8.019e-06)
- Two variables are insignificant

$ttable

	Estimate	SE	t.value	p.value
ar1	0.7001	0.0346	20.2122	0.0000
ma1	0.2075	0.0465	4.4582	0.0000
intercept	-60012.6769	11867.0863	-5.0571	0.0000
ambient_temperature	-432.6182	373.7633	-1.1575	0.2474
wind_direction	-18.7551	16.3386	-1.1479	0.2513
wind_speed	27984.5680	576.9068	48.5080	0.0000

1. **arima(1,0,1) deleting two insignificant variables**

- RMSE: 16073.93
- Test for white noise (p-value = 6.078e-06)
- All variables are significant

$ttable

	Estimate	SE	t.value	p.value
ar1	0.7011	0.0341	20.5423	0
ma1	0.2035	0.0463	4.3908	0
intercept	-75857.8216	4098.9101	-18.5068	0
wind_speed	27932.7931	574.5300	48.6185	0

Models Processing

	Estimate	SE	t.value	p.value
ar1	0.8594	0.0373	23.0210	0.0000
ar2	-0.1616	0.0459	-3.5247	0.0004
ar3	-0.0171	0.0363	-0.4704	0.6382
ma1	-0.9605	0.0132	-72.9282	0.0000
wind_speed	27317.2767	588.7599	46.3980	0.0000

4. **arima(1,1,1) < plot shows not stationary**
   - RMSE: 17095.26
   - Test for white noise (p-value = 6.078e-06)
   - All variables are significant

5. **arima(2,1,1) Final model**
   - RMSE: 16052.45
   - Test for white noise (p-value = 0.0003047)
   - All variables are significant

6. **arima(3,1,1)**
   - RMSE: 16050.29
   - Test for white noise (p-value = 0.0002662)
   - Ar3 is not significant

7. **arima(2,1,2)**
   - RMSE: 16059.29
   - Test for white noise (p-value = 0.001501)
   - Ar1 & ar2 is not significant

	Estimate	SE	t.value	p.value
ar1	0.3350	0.2129	1.5731	0.1161
ar2	0.2391	0.1734	1.3785	0.1684
ma1	-0.4383	0.2009	-2.1819	0.0294
ma2	-0.5143	0.1979	-2.5992	0.0095
wind_speed	27315.5935	589.4691	46.3393	0.0000

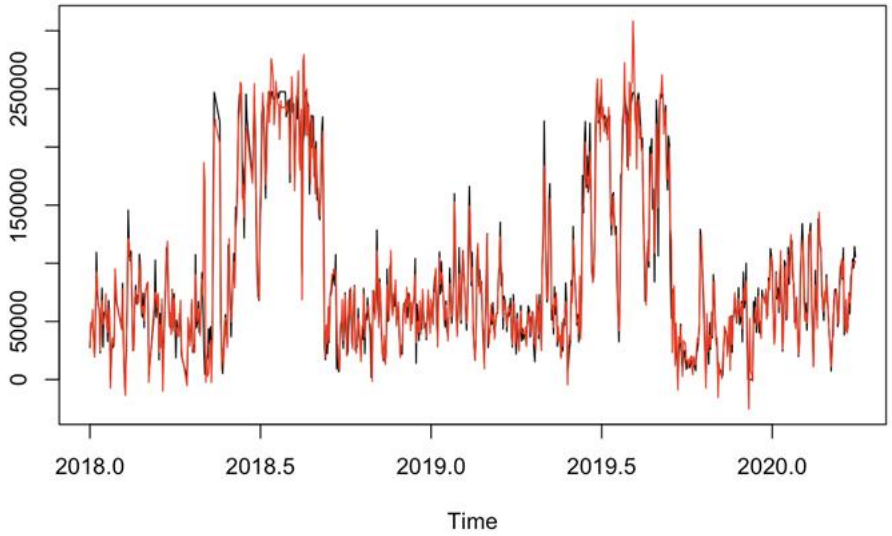
Model Comparison

	Model	RMSE	MAE	Issue involved
(1)	Auto.Arima(1,0,1) + temp + direction + speed	16046.99	10326.17	Residuals are not white noise; temperature and direction are not significant
(2)	Arima(1,0,1) + speed	16073.93	10296.59	Residuals are not white noise; relatively high forecast error
(3)	Arima(1,1,1) + speed	17095.26	11094.76	Residuals are not white noise; high forecast error
(4)	Arima(3,1,1) + speed	16050.29	10314.68	Residuals are not white noise; ar3 is not significant
(5)	Arima(2,1,2) + speed	16059.29	10298.89	Residuals are not white noise; ar1 and ar2 are not significant
(6)	Arima(2,1,1) + speed	16052.45	10319.95	Residuals are not white noise

Final Model - Arima (2, 1, 1) + Speed

```
fit3_AR <- sarima(turbine_tsi, 2, 1, 1,xreg = day_turbine1$wind_speed)
fit3_AR
```

```
$ttable
       Estimate        SE  t.value p.value
ar1      0.8637    0.0361  23.8942       0
ar2     -0.1753    0.0355  -4.9422       0
ma1     -0.9621    0.0125 -76.8522       0
xreg 27316.7580 589.2288  46.3602       0
```



All the terms are significant, and the generated forecast fits well with the actual data line.

Final Model - Arima (2, 1, 1) + Speed

```
fit_AR3 <- Arima(turbine_tsi, xreg = day_turbine1$wind_speed, order = c(2,1,1))
summary(fit_AR3)
checkresiduals(fit_AR3)
```

```
Regression with ARIMA(2,1,1) errors

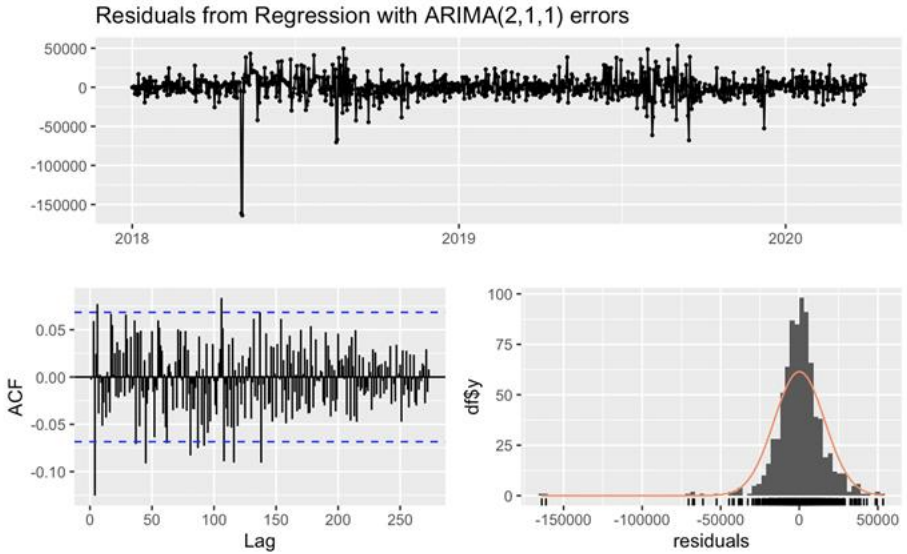
Coefficients:
         ar1      ar2      ma1        xreg
      0.8637  -0.1753  -0.9621  27316.7580
s.e.  0.0361   0.0355   0.0125    589.2288

sigma^2 = 259262026:  log likelihood = -9094.11
AIC=18198.21   AICc=18198.29   BIC=18221.75

Training set error measures:
                    ME     RMSE      MAE  MPE MAPE      MASE          ACF1
Training set 95.09323 16052.45 10319.95 -Inf  Inf 0.2373695 -0.002897159

        Ljung-Box test

data:  Residuals from Regression with ARIMA(2,1,1) errors
Q* = 229.75, df = 161, p-value = 0.0003047
```



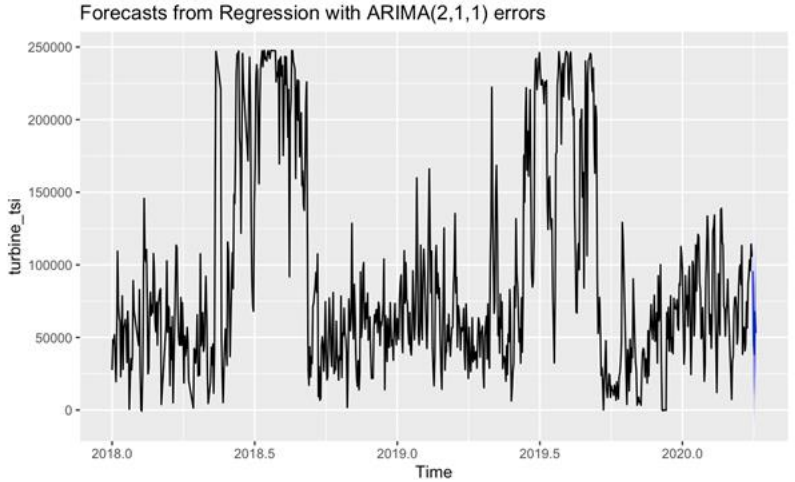
The forecast error is within an acceptable range, but the residuals are not white noise.

Forecast Result - Arima (2, 1, 1) + Speed

```
## Create matrix of covariates for next 5 time periods
```{r}
xdat <- c(5.73, 4.03, 3.88, 5.01, 4.51)
xdat1 <- matrix(xdat, nrow = 5, ncol = 1, byrow = TRUE)

xregmat = day_turbine1$wind_speed
```
```

```
##Re-run model with ARIMA and produce forcast for next 5 days
```{r}
fit_AR3_v2 <- Arima(turbine_tsi, xreg = xregmat, order = c(2,1,1))
forecast(fit_AR3_v2, xreg = xdat1, h = 5)
autoplot(forecast(fit_AR3_v2, xreg = xdat1, h = 5))
```
```



The daily forecast for Active Power generation for turbine for the next 5 day period

Date	Forecast(kW)
3/31/2020	95691.92
4/01/2020	45254.71
4/02/2020	38432.54
4/03/2020	67648.10
4/04/2020	53040.16

Thank you!